| | |
|---|---|
| **Iniziato** | mercoledì, 1 febbraio 2023, 16:37 |
| **Stato** | Completato |
| **Terminato** | mercoledì, 1 febbraio 2023, 16:50 |
| **Tempo impiegato** | 13 min. 25 secondi |
| **Valutazione** | **8,00** su un massimo di 9,00 (**89**%) |

**Domanda 1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Si prendano in esame due dataset `df1` e `df2`: ognuno dei dataset contiene i valori campionari di due variabili, rispettivamente $(x1, y1)$ e $(x2, y2)$. Nella tabella qui sotto vengono mostrati alcuni indici di sintesi dei campioni:

| **x1** | **y1** | **x2** | **y2** |
|---|---|---|---|
| Min. :2.05 | Min. : 4.62 | Min. :1.17 | Min. : 3.63 |
| 1st Qu.:2.45 | 1st Qu.: 7.59 | 1st Qu.:1.98 | 1st Qu.: 6.89 |
| Median :2.89 | Median : 8.71 | Median :2.85 | Median : 8.77 |
| Mean :2.93 | Mean : 8.85 | Mean :2.93 | Mean : 8.85 |
| 3rd Qu.:3.41 | 3rd Qu.:10.02 | 3rd Qu.:3.89 | 3rd Qu.:10.67 |
| Max. :3.90 | Max. :13.05 | Max. :4.87 | Max. :14.58 |

Due modelli lineari semplici vengono stimati: nel modello A y1 è la variabile risposta e x1 è la variabile esplicativa, mentre nel modello B y2 è la variabile risposta e x2 è la variabile esplicativa. Le stime di intercetta e coefficiente angolare del modello A sono:

```
(Intercept)          x1
       2.94        2.02
```

Sapendo che il coefficiente angolare del modello B è 2.02 si indichi il valore dell'intercetta del modello B:

Risposta: 2,94 ✔

La risposta corretta è : 2,94

**Domanda 2**

Risposta corretta

Punteggio ottenuto 4,00 su 4,00

La tabella mostra l'output di R dopo che si è stimato un modello lineare per una variabile risposta Y e una variabile esplicativa X:

```
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)    0.999      2.544   0.393   0.6955
 x              1.521      0.601   2.530   0.0134
```

Si indichino le risposte corrette:

a. ○ La correlazione tra X e Y è negativa.

◉ La correlazione tra X e Y è positiva. ✔

Punteggio ottenuto 1,00 su 1,00

La risposta corretta è: La correlazione tra X e Y è positiva.

b. ◉ Si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 0 al livello di significatività del 5%. ✔

○ Non si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 0 al livello di significatività del 5%.

Punteggio ottenuto 1,00 su 1,00

La risposta corretta è: Si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 0 al livello di significatività del 5%.

c. ○ Si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 0 al livello di significatività del 1%.

◉ Non si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 0 al livello di significatività del 1%. ✔

Punteggio ottenuto 1,00 su 1,00

La risposta corretta è: Non si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 0 al livello di significatività del 1%.

d. ○ Si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 1 al livello di significatività del 5%.

◉ Non si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 1 al livello di significatività del 5%. ✔

Punteggio ottenuto 1,00 su 1,00

La risposta corretta è: Non si può rigettare l'ipotesi nulla che il valore del coefficiente angolare sia 1 al livello di significatività del 5%.

La correlazione tra X e Y è positiva: la stima del coefficiente angolare è positiva, indice che la correlazione campionaria tra x e y è positiva.

Dall'output possiamo notare che il p-value del test $H_0 : \beta_1 = 0$ VS $H_1 : \beta_1 \neq 0$ è 0.013, quindi si può rigettare $H_0$ usando un livello di significatività del 5% ma non ad un livello dell'1%.

Quando cambiano ipotesi nulla ed alternativa e il sistema di ipotesi è $H_0 : \beta_1 = 1$ VS $H_1 : \beta_1 \neq 1$ la statistica test diventa $(\hat{\beta}_1 - 1)/se(\hat{\beta}_1) = 0.867$. Questo valore è piccolo rispetto alle code di una T (per qualunque numero di gradi di libertà) e non si può quindi rifiutare $H_0$ al 5%.

a. FALSE. / TRUE.
b. TRUE. / FALSE.
c. FALSE. / TRUE.
d. FALSE. / TRUE.

**Domanda 3**

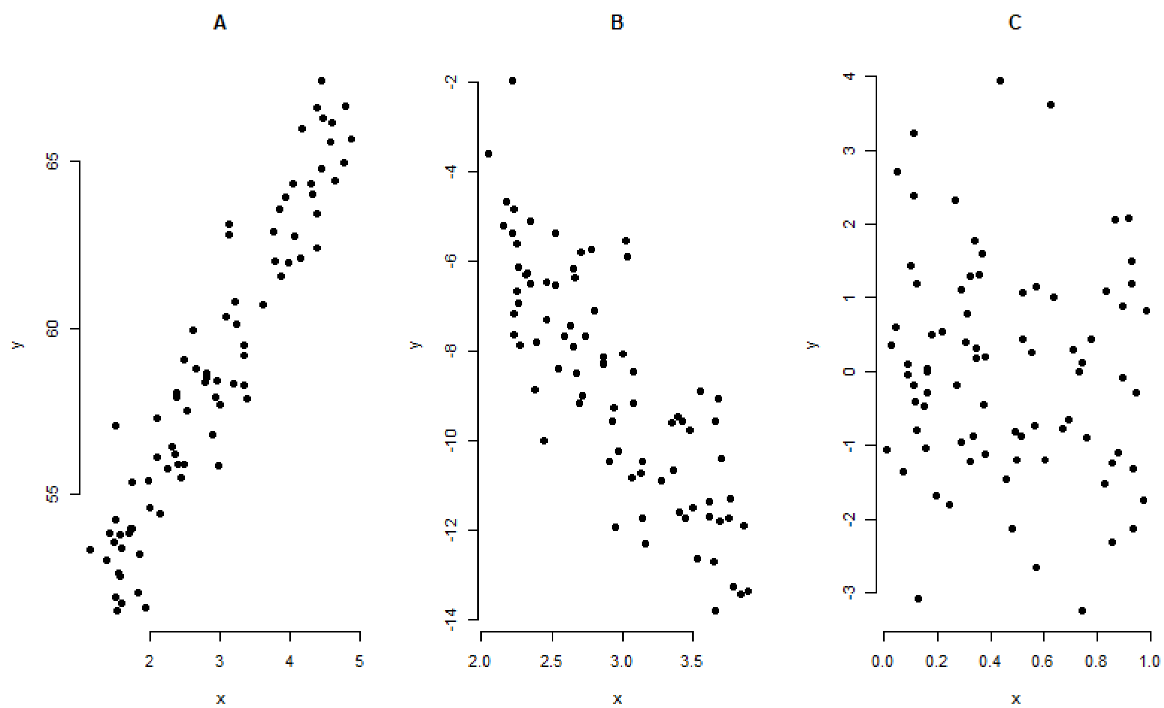
Risposta corretta

Punteggio ottenuto 1,00 su 1,00

La tabella mostra l'output di R dopo che si è stimato un modello lineare per una variabile risposta Y e una variabile esplicativa X:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.94      0.917    3.21 1.94e-03
x1             -3.98      0.308  -12.93 4.27e-21
```

Si indichi quale dei tre grafici è il grafico che mostra i campioni $x$ e $y$ usati per stimare il modello presentato sopra:




Scegli un'alternativa:

- a. Grafico A
- b. Grafico B ✔ TRUE
- c. Grafico C

Dato che la stima di $\beta_1$ è negativa il rapporto tra X e Y deve essere inverso: questo esclude il grafico A. La relazione inoltre è forte: la statistica test è molto grande in valore assoluto: si può rigettare l'ipotesi nulla che $\beta_1$ si 0 a livelli di significatività molto bassi. Questo esclude il grafico C, in cui la relazione tra le due variabili è piuttosto debole.

a. FALSE
b. TRUE
c. FALSE

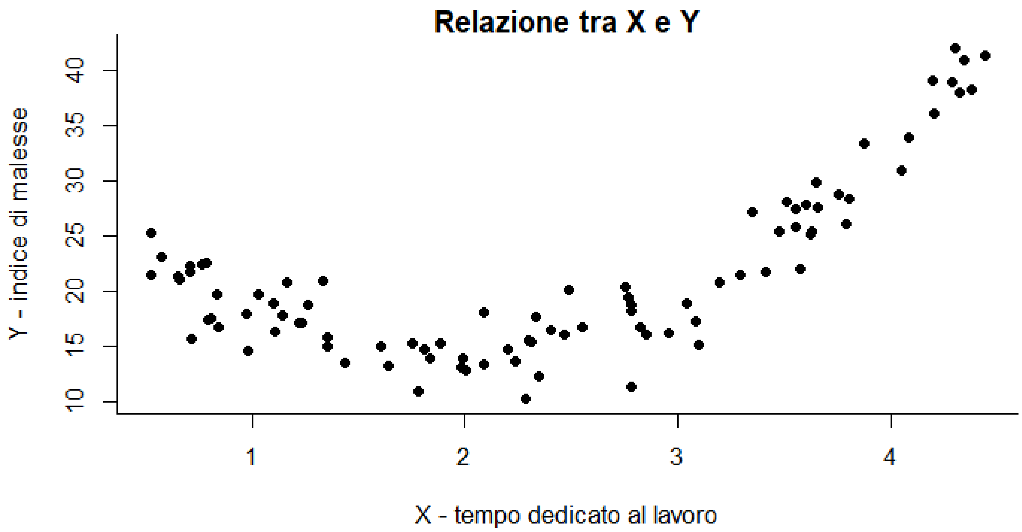
La risposta corretta è: Grafico B

**Domanda 4**

Parzialmente corretta

Punteggio ottenuto 2,00 su 3,00

Un sociologo desidera indagare come il tempo dedicato al lavoro influisca sul benessere fisico e mentale: in un campione di 90 persone tra i 35 e i 50 anni misura un indice che indica il tempo dedicato al lavoro (X) e un indice che misura se la persona mostra segni di malessere (Y). La relazione tra X e Y nel campione analizzato è mostrata nel grafico:



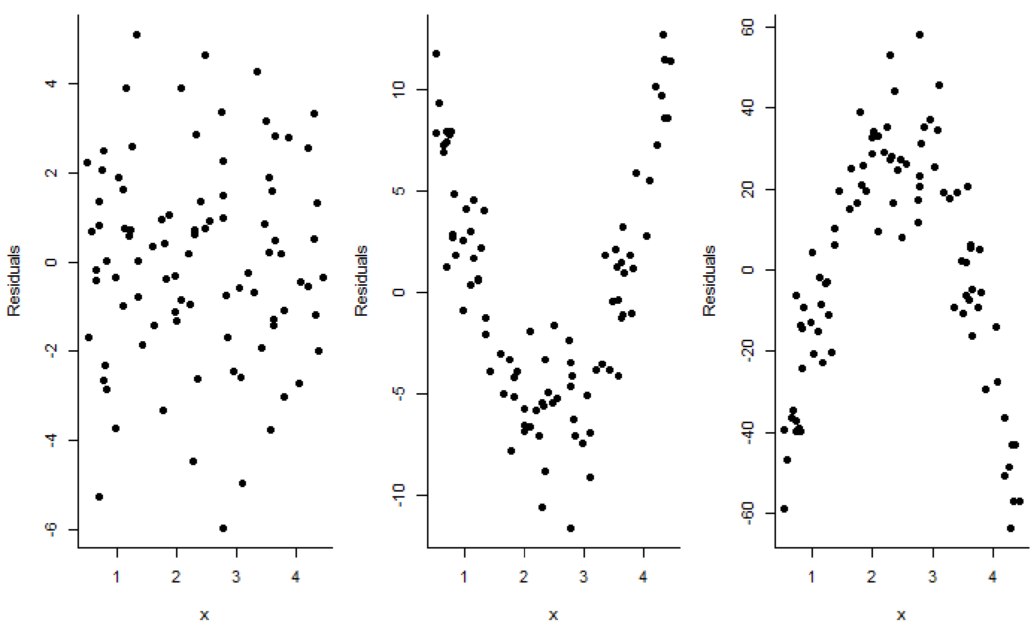

In prima istanza lo studioso stima un modello lineare semplice:

```
fitA <- lm(y~x, data = df)
```

facendo poi un grafico dei residui:

```
plot(x, residuals(fitA))
```

La Figura mostra vari grafici di residui: si indichi quale dei grafici è quello che mostra i residui del modello `fitA`.




Infine si desidera stimare usando il modello `fitA` l'indice di malessere (Y) per un individuo per cui l'indice di tempo dedicato al lavoro è pari a 3.

```
coef(fitA)
```

```
(Intercept)           x
      11.43        4.15
```

```
nd <- data.frame(x = 3)
pp <- predict(fitA, newdata = nd)
```

```
pp
```

```
[1] NA
```

Si indichi inoltre quale modello è probabile possa dare una stima migliore della relazione tra X e Y (cioè aumentare la bontà di adattamento del modello).

a. ○ I residui del modello fitA sono mostrati nel pannello A.

◉ I residui del modello fitA sono mostrati nel pannello B. ✔

○ I residui del modello fitA sono mostrati nel pannello C.

Punteggio ottenuto 1,00 su 1,00

La risposta corretta è: I residui del modello fitA sono mostrati nel pannello B.

b. ○ lm(y

○ sqrt(x)).

◉ lm(y ✖

○ I(1/x)).

○ lm(y

○ x+I(x^2)).

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: lm(y

c. Si indichi il valore mancante nell'output della funzione predict.

23,88

✔

La relazione tra X e Y è quadratica, quindi un modello in cui la relazione tra X e Y viene descritta usando una relazione lineare produrrà residui più grandi e positivi per valori grandi e piccoli di X e valori negativi per la parte centrale del campione di x. I residui del modello fitA sono mostrati nel pannello B.

Di conseguenza è probabile che usare un modello in cui la relazione tra X e Y è quadratica probabilmente migliorerà la bontà di adattamento del modello.

Per un modello lineare il valore stimato per la variabile risposta dato un determinato valore del predittore X si trova con:

$$(\hat{y} \mid x = x^*) = \hat{\beta}_0 + \hat{\beta}_1 * x^*$$

quindi

$$(\hat{y} \mid x = 3) = 11.4 + 4.15 * 3 = 23.85$$

a. Falso / Vero / Falso
b. Falso / Falso / Vero
c. 23.88

◄ Esercizi R

Vai a...

Quiz MLR ►